DØ experience with Monte Carlos in Run 1 and expectations for Run 2

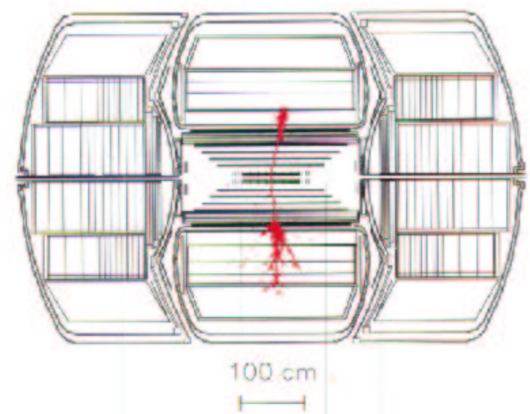
John Womersley
Fermilab

Workshop on Monte Carlo Generator Physics
for Run II at the Tevatron
Fermilab, April 18-20, 2001

Run 1 detector simulation

- DO was an early and pioneering user of GEANT3 in a hadron collider environment
  - e.g. full simulation of missing E_T for detector design studies in 1986
  - Fortran 77 + ZEBRA used throughout D0 code
  - our own interface to event generators: started with ISAJET in early 1980’s

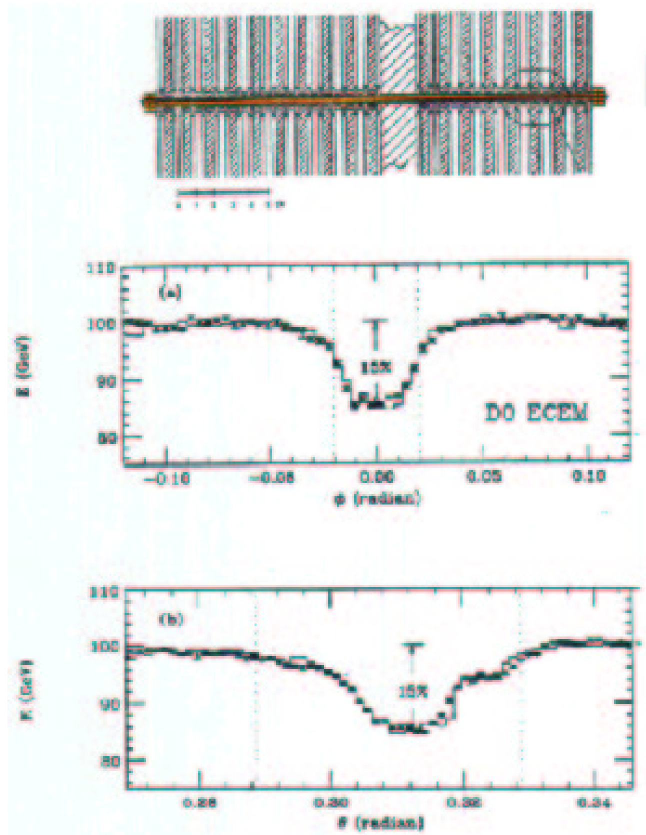
- GEANT3 was an appropriate tool for LEP; we were pushing the envelope in trying to use it for $\bar{p}p$ at 2 TeV
  - one hour per event on a microvax II in 1986



- Three levels of detail implemented in calorimeter
  - full “plate level” simulation
  - “mixture level” simulation
  - shower library

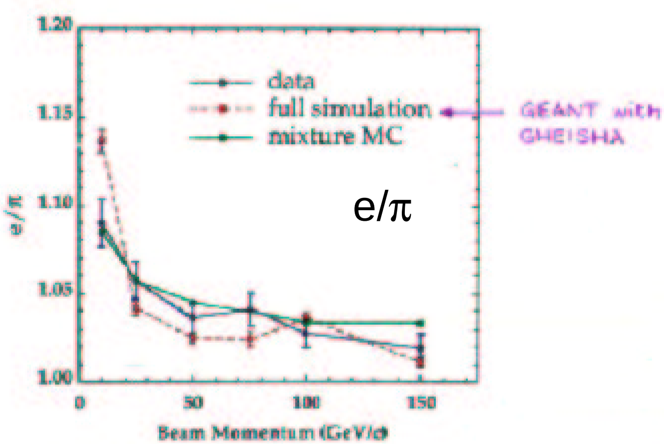
Detailed simulations: EM

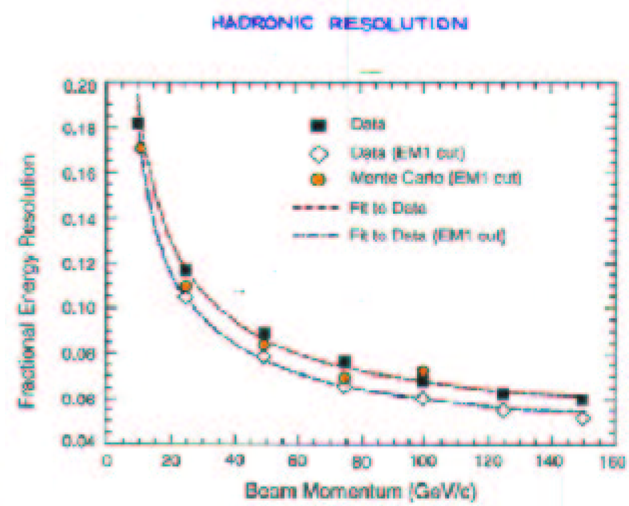
- Exquisite level of agreement achieved between calorimeter testbeam data and full simulation, for EM showers
  - Electron beam scanned across one of the tie-rods in the EM calorimeter



Detailed simulation: hadrons

- The situation for hadronic showers was always a little harder to understand
  - choice of showering programs (GEISHA, FLUKA, etc) and need to define tracking cutoffs
  - 10% level discrepancy in response seen in central hadronic calorimeter between MC and testbeam (normalized to EM); endcap hadronic calorimeter well simulated





Production simulation

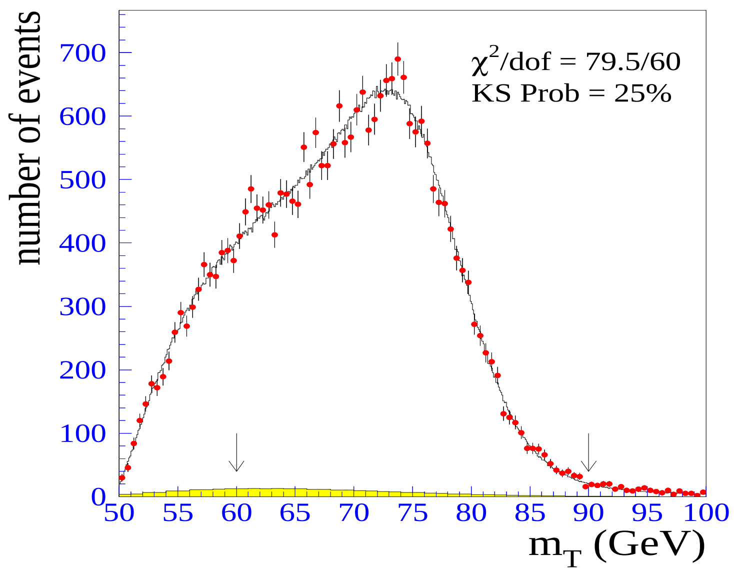
- The CPU time for detailed simulation was always excessive; we essentially never ran this way, except to derive tuning inputs and for one jet corrections study in 1998
- Production running used mixture level simulation and shower library
  - resolution and e/h tuned to match detailed simulation
- Non-calorimetric detectors were less well modelled

Limitations

- Even running in “mixture level”, CPU time was always a concern
  - we never really had enough Monte Carlo
  - we used approximate techniques (shower library)
- So we never developed a great deal of trust in our MC
- Reliance (over-reliance?) on data to make corrections and derive efficiencies

Fast simulations

- To complement GEANT, we developed an array of ad hoc fast simulations for cases where high statistics were a necessity
  - W mass
  - SUSY parameter scans
- Mainly tuned to data



Monte Carlo jet energy scale

- When we derived our jet energy scale, we did it entirely from data
  - photon + jet balancing
  - resolution from dijet events
- At least some of us have developed a bit more confidence in the MC approach since then
  - e.g. reliance on MC for the k_T jet E-scale 1999-2000
- HERWIG+GEANT MC doesn't do too bad a job of describing the data
  - jet resolutions well modelled
  - jet shapes and details of low-E response still hard to understand
    - low E hadrons, noise, and underlying event hard to disentangle

Generators

- We never devoted a great deal of effort to understanding generators:
  - how to tune them
  - what their shortcomings are
  - what are the systematic errors associated with each of them

- Typically might just run PYTHIA and HERWIG with default settings and then compare (a similar approach to that used for PDF's)
  - top mass used ISAJET as a limit to how "different" things could be

Systematic error on m_t	Energy scale	± 4.0
	Generator	± 4.1
	Other	± 2.2
Resulting m_t (GeV/c^2)	$173.3 \pm 5.6(\text{stat}) \pm 6.2(\text{syst})$	

We estimate the uncertainties in modeling of QCD by substituting the ISAJET MC generator [13] for HERWIG, independently for top MC and for VECBOS fragmentation, by changing the VECBOS QCD scale from jet $\langle p_T \rangle^2$ to M_W^2, and by varying the amount of initial and final state gluon radiation in the top MC. The resulting systematic

What's new in Run 2?

- Still using GEANT3 detector description, wrapped in C++, with digitization done in C++
- Continue with plate and mixture calorimeter simulation options, but no shower library (no longer needed for speed)
- Cleaner event generator interface
- Added interface to unified, modularized fast simulation
- The biggest change is in sheer availability of CPU
  - Can now generate ~ 1 million events per day and store in central repository at Fermilab (SAM) via network or tape transfer
    - similar rate to real data.
  - Invites a change in the way we do analyses.
    - But not yet really thought through the implications...
      - e.g. cost of tapes >> cost of CPU!

DØ Monte Carlo Production

- Plan to generate **ALL** MC events off-site:
  - Currently 1 CPU can fully simulate and reconstruct ~500-1000 events/day (3 min/event)
  - Current DØ computational “Grid” ~500 CPU’s
  - Generate 50-100M events/year.

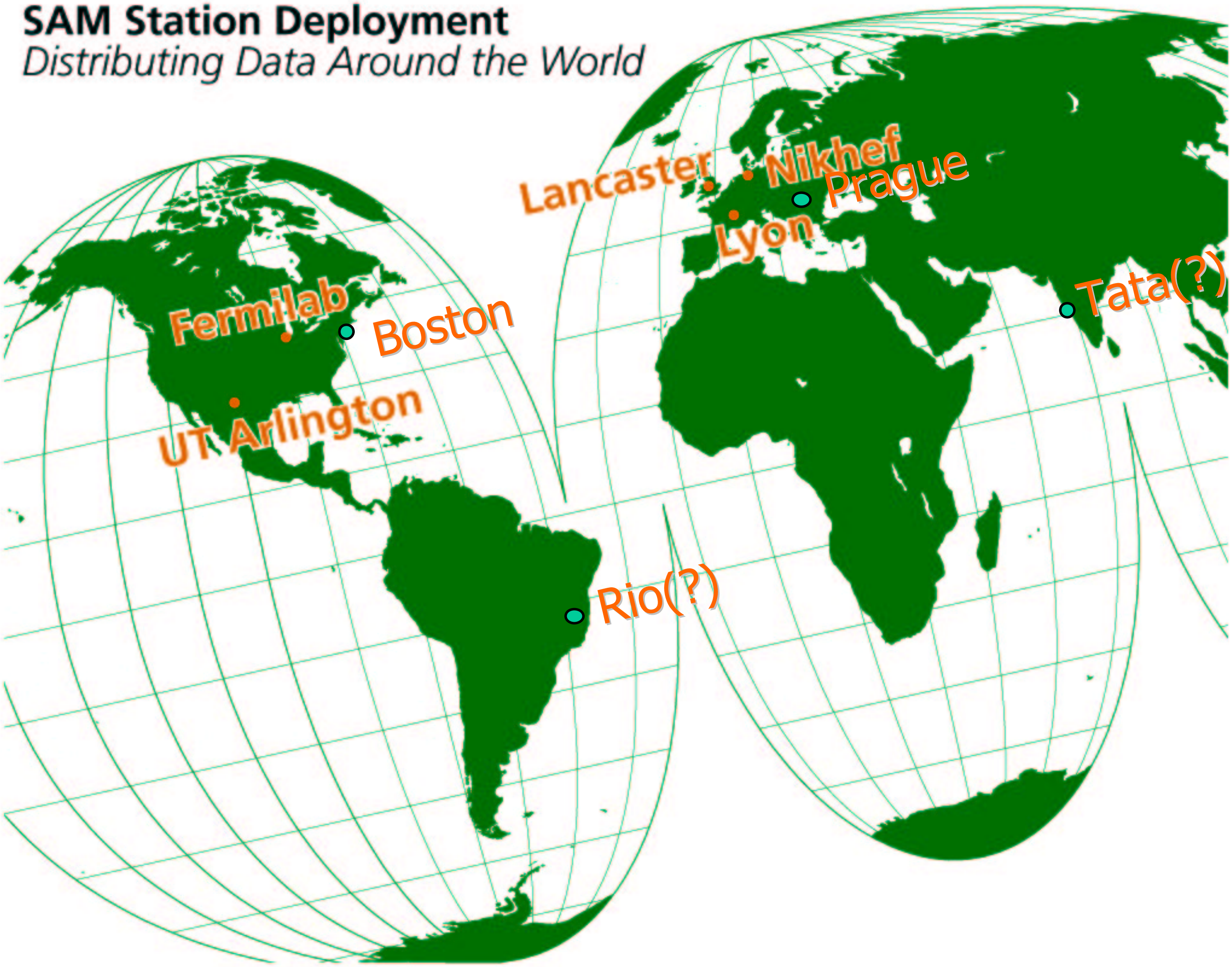
Location	# CPU’s
NIKHEF	100
U. of Texas (Arlington)	64
Lyon (CCIN2P3)*	100
Boston*	O2000 (192)
Prague (Charles U.)	32
Lancaster	200
Rio	100 (proposed)

*Not Completely DØ

Total Bandwidth to Fermilab ~ 5Mb/sec

- Some farms will be upgraded substantially this year/next year

DØ Production “Grid” Locations



Open issues in generators

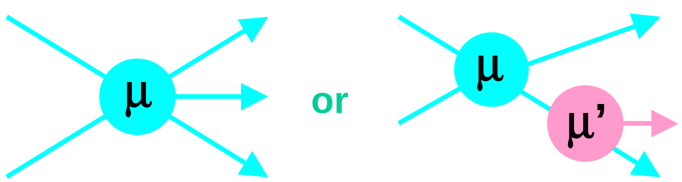
- For jet energy scale, big uncertainties to do with underlying event modelling.
  - How well is this done?
  - How can we improve?

- Minimum bias events (multiple interactions per crossing)
  - understand effects on missing E_T
  - particle multiplicities and energy flow (isolation, pattern recognition)

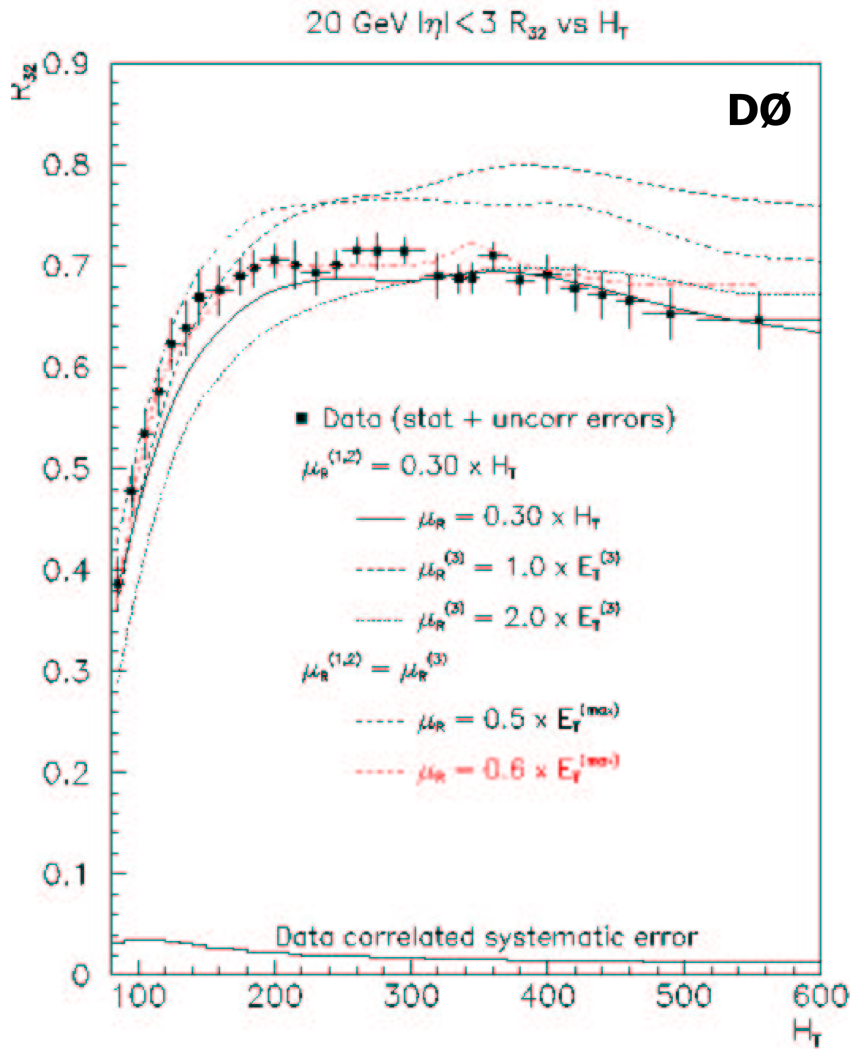
- Hadronization effects in jets
  - shift energies by O(1 GeV) particle vs. parton
  - May be important for top mass measurements in Run II; already shift jet cross sections by 10-20%.

Ratio of 3-jet/2-jet events at DØ

- Ratio predicted reasonably well even by JETRAD (a leading order prediction of R_{32})
- Can any information be extracted on the best renormalization scale for the emission of the third jet?



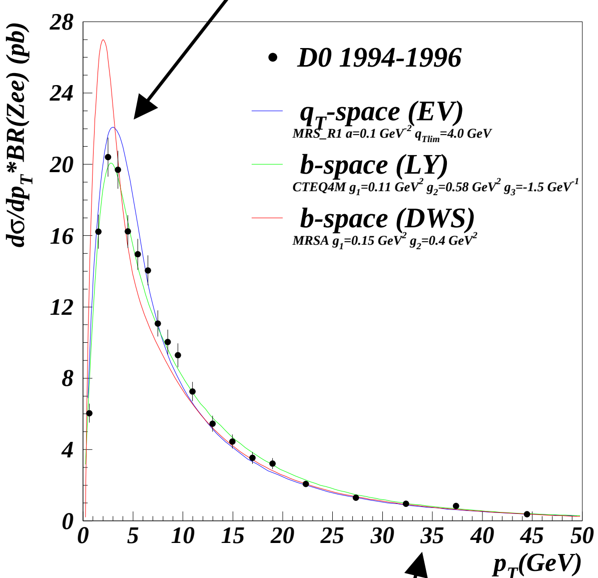
  - Same scale as the first two jets seems better than a scale tied to E_{T3}
  - $\mu = 0.6\ E_T^{max}$ is pretty good
  - $\mu = 0.3\ H_T$ is best as $E_{T3}\uparrow$



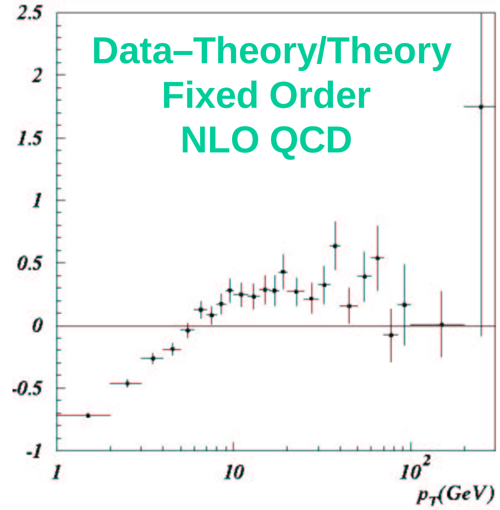
DØ p_T^Z measurement

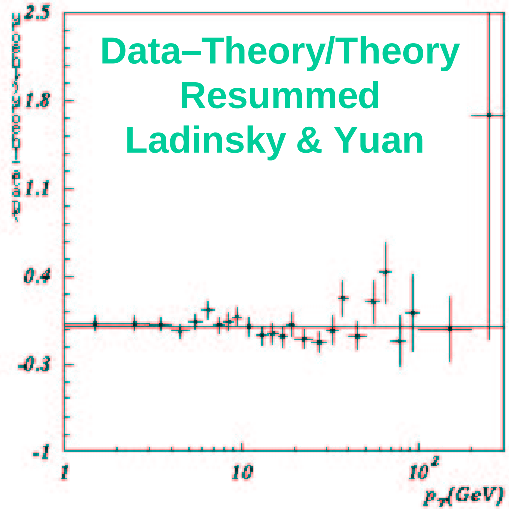
- Phys. Rev. D61, 032004 (2000)

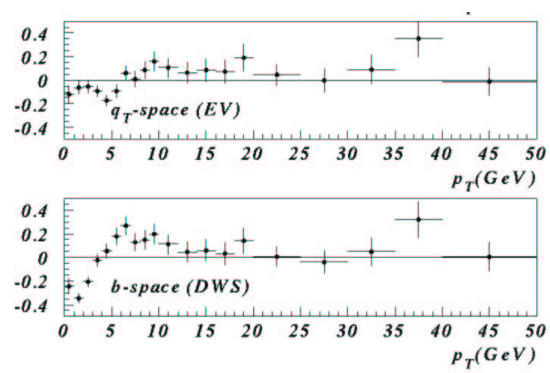
Low p_T (< 10 GeV)
resum large logarithms of m_W^2/p_T^2 and include nonperturbative parameters extracted from the data



Large p_T (> 30 GeV)
perturbative calculation





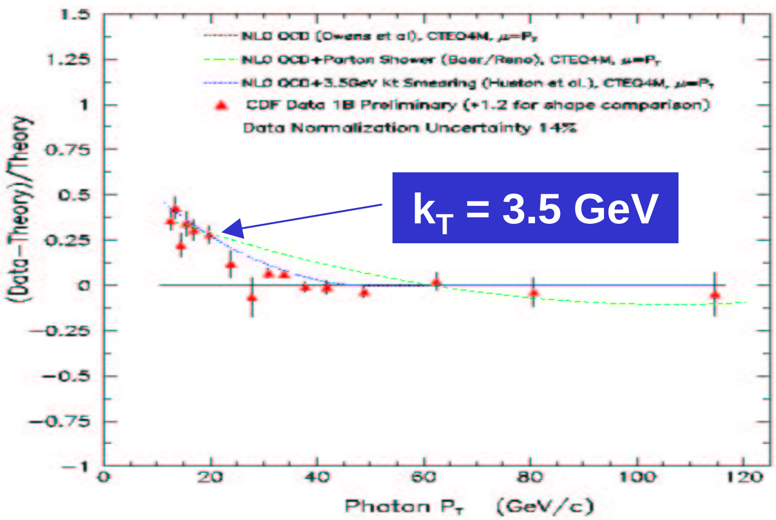


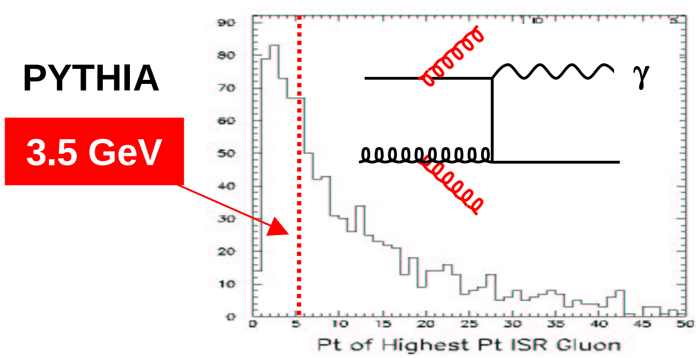
Ellis & Veseli and Davies, Webber & Stirling (Resummed)

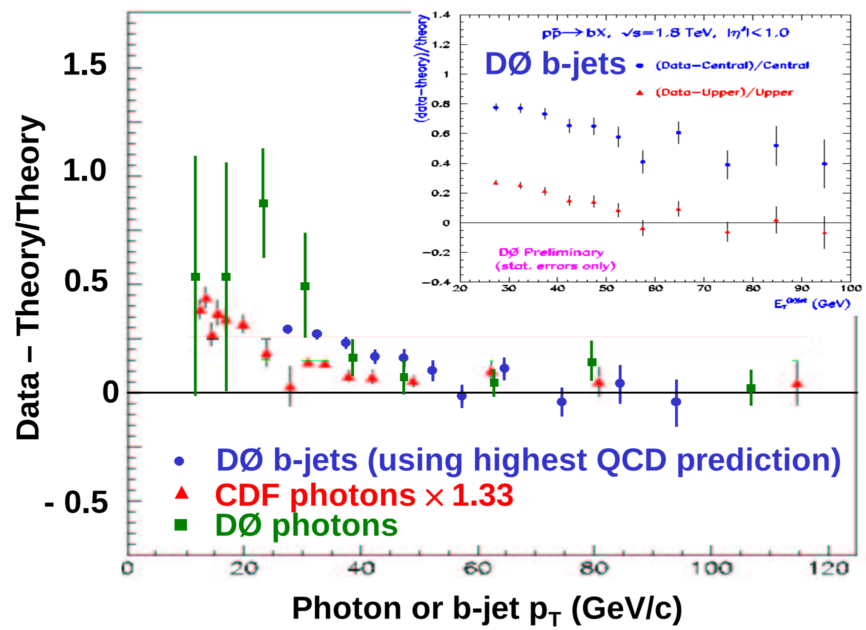
not quite as good a description of the data

Low E_T rise in cross sections

- "k_T" from soft gluon emission







Same effect in b-jets?

The challenge for Run 2

- Effectively exploit the dramatic increase in statistics available
  - Monte Carlo as well as data!
- Do not allow event modeling uncertainties to limit our physics
- Make use of the data we take to reduce these uncertainties
  - perturbative QCD calculations
  - production models
  - PDF’s
  - fragmentation
  - underlying event and minimum bias
  - . . .

Will require an ongoing, open dialog between the experimenters and the phenomenologists: hopefully this workshop can be start of such a process

An observation

- Event simulation is the link between experiment and theory: it is the only way to test data against predictions

- But, despite its importance, it often seems rather neglected
  - "not really experiment"
  - "not really theory"
  - no jobs, no future for the practitioners

- How can we improve the situation?
  - changes in our institutional structures?
  - changes in the way experiments are organized?
  - Initiatives like CTEQ, Physics Frontier Centers?